

Although automatic data analysis methods have long become an irreplaceable norm, visualisation of complex data allows human users/operators to instinctively find interesting trends and novel information, which are too abstract or complex for computational approaches to identify. Scientific discovery games have shown the benefit of visualising complex data where even non-expert human users can uncover hidden trends and surprising conclusions. The proof of human knowledge aided by interactive visualisation triumphing over purely computational analysis methods is best exemplified by the Foldit project [2]. The endeavour involved human players interacting with protein structures visualised in 3D, to find feasible configurations of enzymes which can facilitate further research. Transforming the research into a visual game proved to be a success as Foldit players only took 10 days to find the 3D structure of the AIDS-causing monkey virus, a task

which stumped bio-engineers for 15 years [3]. There is a significant body of research showing that visualisation of data can make a huge impact on solving complex tasks ranging from interplanetary spacecraft trajectory design [4], mapping the brain [5], to DNA and RNA sequence alignment [6], [7].

Similarly to citizen science games, engineering projects can greatly benefit from visualising data in a useful manner, and allowing manipulating the data through an interactive interface. Parallel Coordinates [8], [9] is one of the most popular and efficient techniques for visualising high-dimensional, multivariate data. Each data attribute is represented in the plot as an axis, and a polyline going through every axis represents each datapoint. An example plot is shown in Fig.1 with datapoints depicted in red.

Multidimensional data is generated in the vast majority of engineering applications, simulations, and design. Parallel Coordinates allow a more insightful, high-level perspective of the content, facilitating understanding beyond that of two- or three-dimensional graphing methods, or raw data alone [10], [11]. The dataset presented in this paper, and shown in Fig.1 is a synthetic dataset composed from two separate, but coordinated, studies on aircraft wing design. The common characteristic between the two sub-datasets is the manufacturing tolerance of the 2D aerodynamic profile. This describes the amount of deviation, in millimetres, from the nominal aerodynamic shape. The first part of the complete dataset concerns aerodynamic qualities considering angle of attack, mach number, and calculating lift-to-drag ratio, wave drag, drag count, and other aerodynamic coefficients. The second study explored the manufacturing aspects of the wing considering costs, different types of manufacturing process, and other manufacturing coefficients. Finally, the dataset includes a value assessment criteria following the methodology that is described in [12], [13].

We present the updated and extended Parallel Coordinates graphing module in the Cambridge Advanced Modeller (CAM) [14]. The module emphasises intuitive and interactive user interface, significantly broadens the perspectives on plotted data, and allows integration and simultaneous use with external software, as part of large engineering projects.

The evolution of display technology brought about a wave of new opportunities for the visualisation and manipulation of big data. Ultra-high resolution and multi-touch screens allows increased control and detailed visual depiction of the designs. With this continuing trend under consideration, CAM's every feature in the extended Parallel Coordinates implementation has been developed to fulfil the requirements of touch-screen and high-definition displays.

Finally, to achieve best performance, we intertwine interactive visualisation with AI techniques to assist the designer in the analysis and facilitate a further understanding of the data via a surrogate model.

The main features of our toolkit introduced in this paper are divided into two sections, and are outlined below.

Interactive Parallel Coordinates:

- Interactive & touch-screen enabled interface for Parallel Coordinates plots.
- Interval selections.
- Free-form polygon selections on Scatter plots.
- Multiple simultaneous selections across Scatter and Parallel Coordinates plots.
- Creating new dimensions via parsed mathematical and logical expressions.

AI-aided Design:

- Kriging-based Surrogate model.
- Generating new data points via Interpolation.
- Dynamic and online predicting of outputs from parameter changes.
- Uncertainty Quantification for interpolated values.

II. RELATED WORK

With the rise of big data in recent past, visual analytics became a growing trend in engineering, computer science, and other domains of technology. Visualising complex multivariate data is an invaluable method for facilitating the understanding of content by human analysts. As a result, there is a great body of research devoted to multivariate data visualisation.

Since the invention of Parallel Coordinates [8], [9], [15], [16] the approach has received substantial research attention and spawned multiple variants which cover a multitude of applications [17].

One of the most common techniques used in tandem with Parallel Coordinates is brushing [18], an interaction technique which allows the user to explicitly select a subset of the visualised data to bring into focus. Brushing has been used in conjunction with Parallel Coordinates in the vast majority of implementations, visualising data in various applications such as Computational Fluid Dynamics [19] or Satellite Imaging [20].

New approaches have also been recently explored to reduce overplotting and data cluttering [21] via smarter higher-order brushing and guiding the user.

More recently, the development of Parallel Coordinates plots have been enhanced to include a time axis, pushing the plots into 3-dimensional space [22]. This extension allows analysis of the evolution of attribute values over time, though the clarity of the plots is significantly reduced.

The various "flavours" of Parallel Coordinates perform well for small- to medium-sized datasets, however big datasets suffer from severe overplotting and clutter issues. There have been some attempts at solving these problems, most notably with clustering the visualised data [23], [24]. Clustering is mostly done a priori, i.e. the data is clustered before being plotted, however the approach is limited by the clustering algorithm. The cluster and density representations in parallel coordinates plots have been further extended to encompass detailed information and display underlying patterns [25], [26], [27].

Andrienko and Andriendko [28] developed a Parallel Coordinates plot approach which mitigates the issue of overplotting,

by distinguishing between subsets (classes) of the data, though the number of classes in the data is limited.

Nguyen and Rosen proposed the Data Scalable Parallel Coordinates (DSPCP) which exploit the point/line duality to show the relationships and their consistency in the data. DSPCPs perform very well identifying linear and non-linear patterns, noise, and outliers in large datasets but do not provide the precision required for engineering design applications.

Brushing, clustering, and increased dimensionality of Parallel Coordinates plots increase the clarity of large plotted dataset, though certain drawbacks persist. Some approaches have been made to reduce cluttering in the plots [29], however, to date there is no definitive method for mitigating the overplotting and cluttering issues of large datasets.

More recently, some implementations of Parallel Coordinates provided users with interactive interfaces. D³ [30] is a large web-based visualisation toolkit which includes Parallel Coordinates plots. However, it only supports basic brushing and single selections. XDAT [31] displays multidimensional data via Parallel Coordinates and scatter plots with emphasis on clarity. It also supports manual clustering of plotted data for pattern discovery. ParallAX [32] is arguably the most extensive data mining tool based on parallel coordinates. It allows the user to select subsets by brushing and pinch selections, as well as display the data as a scatter plot and use polygon selections. ParallAX also includes an algorithm for data classification.

However, all of the aforementioned approaches and tools lack advanced machine learning methods for in-depth data analysis, and few give the user sufficient freedom to tackle the arising issues and maintain the accuracy required for engineering design projects.

III. INTERACTIVE VISUALISATION IN CAM

Visualising complex multivariate data improves the understanding but it still runs the risk of overplotting and clutter. The Cambridge Advanced Modeller implemented basic Parallel Coordinates plot features though it lacked advanced functionality to sufficiently manipulate plane design data. To efficiently and intuitively manipulate the visualised data we extended the functionality of CAM's Parallel Coordinates with a set of interactive and touch-screen friendly features. Throughout the design cycle, we continuously work on streamlining the GUI to reduce complexity and enable a more intuitive touch-screen experience (i.e. reducing the number of clicks, simplifying the control flow, etc.). Note that while all presented features have been designed with touch-screen devices in mind, they can be just as easily operated via a traditional mouse/keyboard interface.

A. Interval Selection

The main interface of CAM's Parallel Coordinates module is an interactive plot with supporting functions. Directly in the graph the user can define interval selections of the datapoints simply by clicking on the axes. Each click defines an upper or lower bound of the selection which will consist of the datapoints which contained in between the bounds. Fig. 2

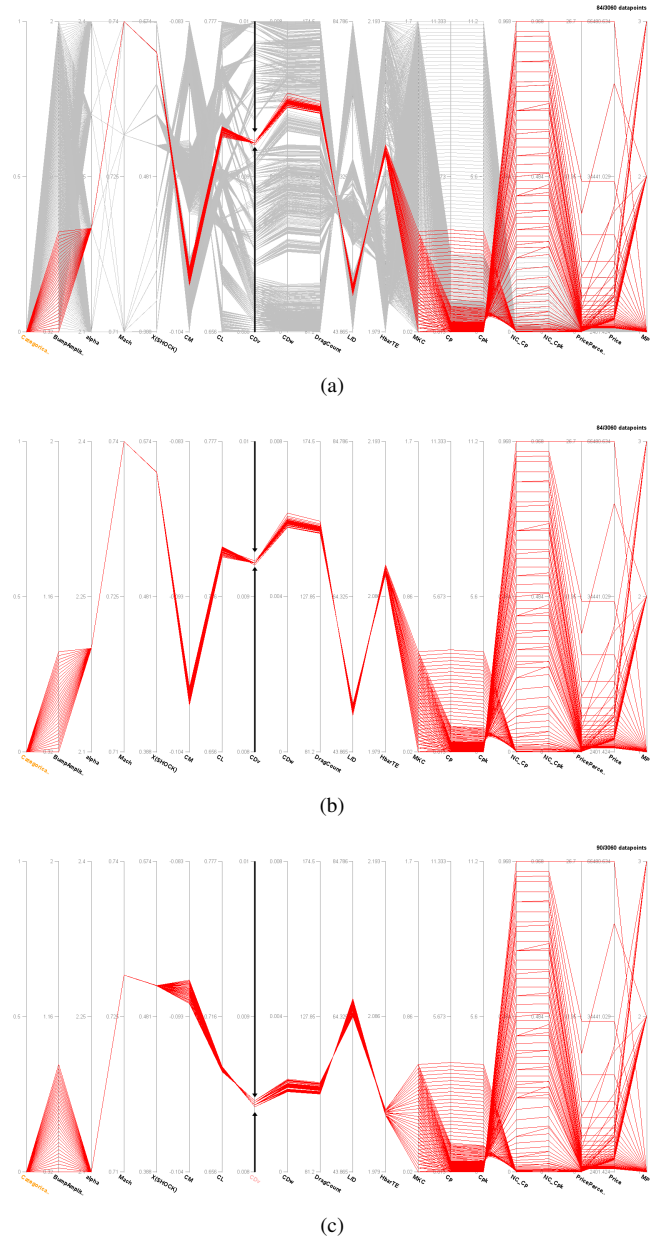


Fig. 2. An example of interval selection showing in (a) and hiding in (b) the unselected datapoints.

depicts a selection made on axes in the plot, where the black arrow tips are the selection bounds. To tackle overplotting and clutter, the user can hide the unselected points (Fig. 2(b)). Note that for the purposes of precision, exact numeric selection bounds can be set via a dialog, if the graphical interface is not sufficiently accurate.

Selecting designs requires considering constraints which prune the infeasible datapoints. Frequently, constraints are defined in relation to one another, thus adjusting constraints on one axis might require a series of adjustments on other axes. To prevent this cascade of parameter tweaking, CAM's Parallel Coordinates allow users to adjust the interval selection bounds on multiple axes simultaneously, while maintaining the proportions and relations. The user selects the appropriate

axes, then slides the selection bounds accordingly by dragging the mouse/finger/stylus. Fig. 2(c) shows the Parallel Coordinates plot from Fig. 2(b) after simultaneously adjusting the selection bounds on the selected three axes (labeled in pink). As a result, a new solution has been found while maintaining the relationship between the constraints on the axes (Fig. 2(c)).

The structure of the data inside the datafile determines the order of the axes in the Parallel Coordinates plot. However, the datafile is not structured to fit the visualisation, rather it is determined by the analytics software. To facilitate clarity of and improve insight into the plotted data, we allow the axes to be reordered to fit the designer's requirements. The designer has freedom to reorder any and all axes, one or multiple simultaneously.

Selecting a design requires satisfying sets of varying constraints. However, defining these sets can be difficult or even impossible on a single Parallel Coordinates plot. To mitigate this issue, we implemented the ability to make multiple sets of interval selections on the plot. In practice, multiple identical Parallel Coordinate plots exists which all can be separately manipulated. Fig. 3(a)-(d) show selections on the same dataset according to different sets of constraints, multiple criteria, and user's preferences. An overview of all the selected datapoints across all interval selections is displayed through superimposition of the datapoints (shown in Fig. 3(e)).

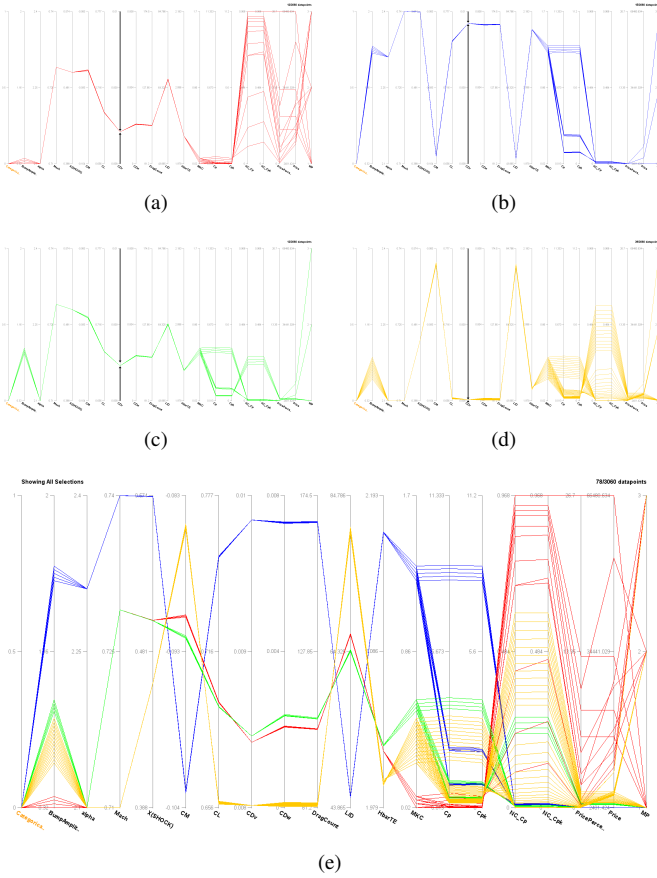


Fig. 3. Separate interval selections on a sample dataset ((a)-(d)) and their superimposition (e).

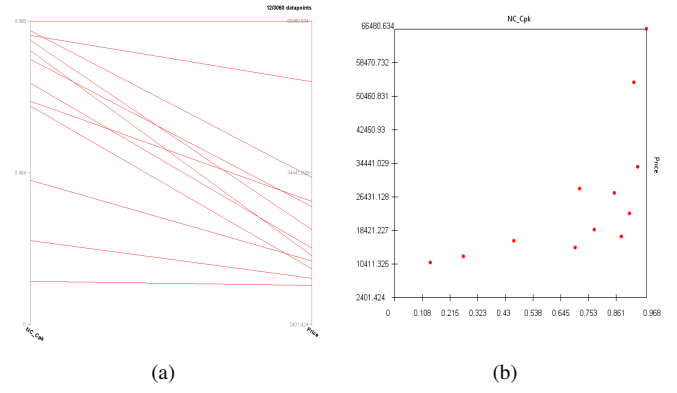


Fig. 4. A simple Parallel Coordinates plot and a corresponding scatter plot representation.

B. Scatter Plots

Scatter Plots [33] provide significant support to visualisation with CAM's Parallel Coordinates. In our implementation, points from any two axes in the Parallel Coordinates plot can be graphed onto an X-Y plane. A simple example is shown in Fig. 4. The Parallel Coordinates data (Fig. 4(a)) and the corresponding representation of the data as a 2D scatter plot (Fig. 4(b)).

Scatter plots allow high-level overview of two-dimensional information which might not be easily understandable by examining the entire Parallel Coordinates plot. To extend the functionality, we implemented data selection features directly on the scatter plots. The user draws selection regions which contain datapoints deemed interesting. The selection regions are polyhedra which correspond to particular two axes (i.e. the selection region only exists between two axes, it will not be visible when looking at different axes).

For simplicity the user clicks to define the position of corners of the region. Once finished, the datapoints contained in the drawn polyhedra are highlighted, while all other datapoints are greyed out. Any changes made in the scatter plot are immediately reflected in the main Parallel Coordinates chart. Fig. 5(a) depicts an example scatter plot selection, while Fig. 5(b) shows the selection's effects on the parallel coordinates plot. In the figure, red points in the scatter plot correspond to the red datapoints in the Parallel Coordinates plot, while the black points outside the scatter plot selections correspond to the greyed-out datapoints.

In order to give the designer enhanced control and additional perspectives on the design data, multiple scatter plots can be in operation simultaneously. This allows the designer to manipulate the data in accordance to varying sets of constraints and decision criteria. However, not all constraints are mutually exclusive, certain projects require consideration of designs which fulfil only part of the constraints.

As a result, we allow for different selections to be expressed in separate scatter plots and the designer can choose how the main Parallel Coordinates plot will reflect the scatter plots selections. Based on the user's preference, the Parallel

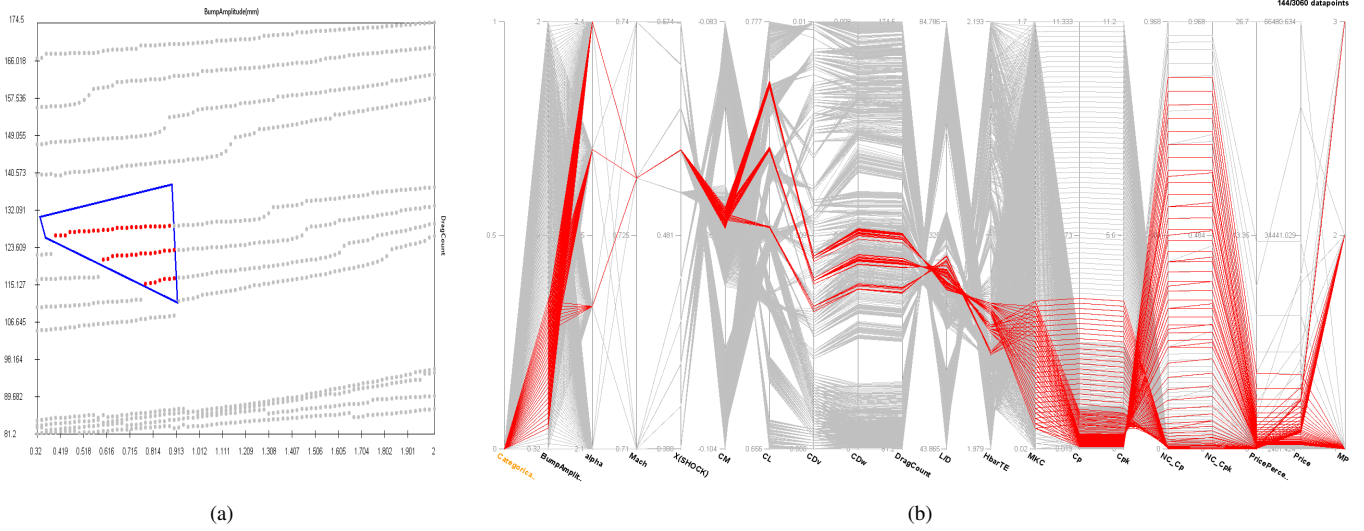


Fig. 5. A sample selection on a scatter plot in (a) and its reflection on the Parallel Coordinates plot in (b). The scatter plot X axis is the BumpAmplitude and the Y axis is DragCount from the Parallel Coordinates plot in (b).

Coordinates plot can either display the *intersection of selections*, i.e. only highlight the datapoints which appear in all made selections, or the *sum of selections*, i.e. displaying all datapoints which are contained in at least one selection. Fig. 6 displays identical sets of selections (an interval selection on the C_{pk} axis in (c/d), and two polygon selections in (a/b)) and shows the difference between the sum/intersection option. Intersection is depicted in figures 6(a) and (c) where only points common to all selections are shown in red, while figures 6(b) and (d) show the sum of selections (i.e. points in at least one selection are highlighted).

C. Inserting New Dimensions

In some cases, the existing data can have the basic building blocks for an in-depth analysis but might not be presented in a clear manner to help the designer. For those instances the designer has to manually carry out the calculations in order to extract the relevant information and make informed design decisions. This detracts from a smooth workflow and might require employing other software or writing functional scripts to process the data. Sometimes, it might be unfeasible to process large datasets and calculate the necessary values for further analysis.

To tackle this issue, we propose a smart approach to add new dimensions to the existing dataset based on the designer's input in the form of a mathematical expression to define an additional value for all existing datapoints. The expression is parsed by mXparser [34], a versatile and fast parser and evaluator for mathematical expressions written in plain text, which allows great freedom in behaviour. With mXparser, our Parallel Coordinates software can add dimensions which are calculated using linear and non-linear functions, conditional and logical statements, derivatives and integrals, iterative and recursive operators, and more.

Building additional points of reference for the displayed dataset is important for the designer as it can reveal hidden patterns and obscured details, possibly with significant impact on the design decisions. Often two parameters can seem uninteresting at first glance but when paired together, can introduce an additional perspective. As an example, consider the lift and drag values for a proposed aircraft wing design, separately, they can both be within given aerodynamic norms. However, new information is added when presented with the lift to drag ratio values which are crucial to assessing a design. Exposing this information in our Parallel Coordinates tool is simply done by adding a new axis which is populated by *lift/drag* values for each datapoint in the displayed set.

D. Tracking Data Change

Visual analytics is used in conjunction with a broad variety of software packages responsible for filtering, standardising, and preparing data for subsequent phases of the design process. Large engineering endeavours are carried out by multiple engineers and/or groups of engineers all of whom have different preferences in terms of design software and methods. This may result in multiple incomplete versions of the data in simultaneous use, as well as wasted effort working with the incorrect data. It is crucial to ensure a seamless flow of information between the different simultaneous developments made in the project. Furthermore, designers are in need of the most up-to-date information. To achieve this requirement, we implemented a tracking feature which notifies the Parallel Coordinates user whenever the currently-active datafile is modified.

Example. Imagine a simulation experiment which optimises a vast number of designs. As each optimisation is completed, the resulting data for that design is written to a datafile. Processing all of the designs can take a significant amount

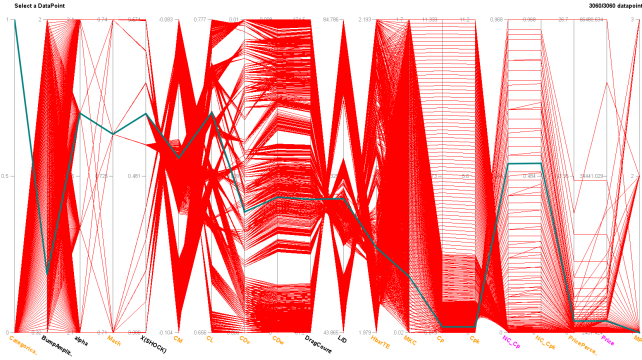


Fig. 8. A selected datapoint with adjusted inputs and Kriging-computed output values.

designer to label the parameters as input/output. Once selected, the designer can then add a new set of parameter values and the Kriging module will interpolate the output values based on the existing dataset.

The Kriging module allows for a rapid prototyping of designs used to confirm or contradict the designer's intuition. Analysing the plotted data can give the designer ideas on improved designs. However, in engineering projects, evaluating these designs can involve expensive and time-consuming simulations. Experimental designs based on designer's intuition are often discarded, thus wasting much computational effort to evaluate them. Instead, our Kriging interpolation method allows for a quick evaluation of new designs based on the existing plotted data. This improves the designer's knowledge and allows for faster analysis of the data and reduces the parameter space for costly simulations. This results in improved accuracy, quicker iterations, and an overall reduced duration of the design analysis.

B. Manipulating Existing Datapoint Values

Additionally, the designer has the choice to manipulate the parameter values of existing datapoints. The output values of the selected datapoint are immediately updated to respond to the designers controls. This functionality further allows the designer to understand the nature of relationships between design parameters and the sensitivity of each parameter to change.

Direct manipulation of an existing datapoint is based on the aforementioned Kriging implementation (similarly to interpolating new designs). After selecting output variables, any datapoint can be selected for modification. Any value dimension of the selected datapoint can be simply dragged across the axis to change its numeric value. Input dimensions can also be coupled for simultaneous and proportional change in values. The output variables' values are dynamically recalculated in response to the invoked changes, based on the computed surrogate model.

Figure 8 depicts a selected datapoint (green polyline) whose input values have been modified, and whose output values have been computed using the Kriging-based interpolation method. All predictions are calculated and displayed dynamically in

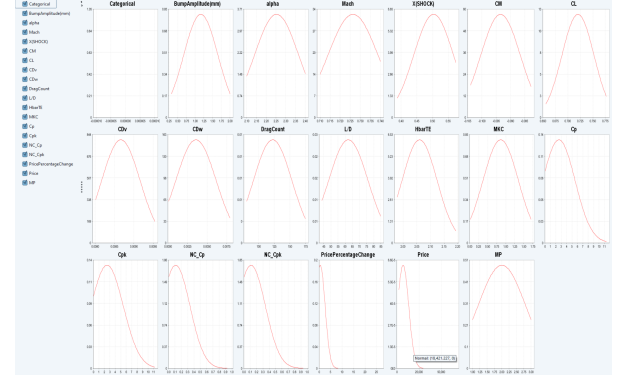


Fig. 9. Dataset Variability: value distributions for selected variables.

response to the input adjustments. Input dimensions are signified by black axis labels, output axis labels are shown in pink/violet, and neutral variables (i.e. omitted from Kriging) have yellow axis labels.

C. Uncertainty Quantification & Management

Engineering Design projects are notoriously difficult and the designers have to account for a wide variety of risks. Faults found during the development phase can be extremely expensive, time-consuming, and difficult to implement. Furthermore, both input and output values have some degree of associated uncertainty. As a result, the design phase has to ensure that the chosen specifications are robust and risks are minimised.

Uncertainty quantification methods are a standard practice used in engineering to analyse and manage design risks. To support the designer in this task, our Parallel Coordinates toolkit is equipped with uncertainty management features which facilitate deeper understanding of the variability of the datasets and how uncertainties are propagated from input values to the outputs.

The initial step is to grasp the trends in the input variables. This allows the designer to better assess the uncertainty stemming from the dataset itself. Engineering data is often obtained from companies using their proprietary software and practices. Its accuracy and robustness can be obscured, thus any subsequent analysis requires accounting for possible uncertainties that might have come from previous processes. A straightforward method is to analyse the distribution of the values for each of the design dimensions.

To aid the designer, we implemented a *Dataset Variability* feature which shows Gaussian distributions for each of the selected variables in the dataset. Figure 9 shows the value distributions of all dimensions from our dataset.

All engineering computational processes incur uncertainties and risks. As a result, it is of paramount importance to compute the uncertainty propagation accurately, and minimise the associated risks as much as possible. For this purpose we implemented the uncertainty propagation methods based on the Kriging surrogate model. The input distribution is represented via Chebyshev nodes, which are spread strategically over the range of the dimension values. They are

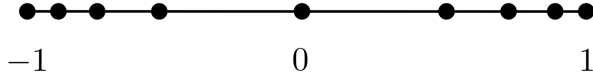


Fig. 10. An example of Chebyshev nodes spread over $[-1, 1]$ interval.

symmetrically placed with nodes at smaller intervals closer to the tails of the distribution, as depicted in fig. 10.

Kriging is used to predict the values of output variables for each selected datapoint. However, the accuracy of the predictions via interpolation are highly dependent on the size and coverage of the dataset. Therefore, the computed output values have an associated probability distribution which shows the output's reliability and accuracy.

The Parallel Coordinates toolkit is equipped with a feature designed to show the input uncertainty for each selected datapoint and propagate these uncertainties to the output variables. The output value distribution is computed using Chebyshev nodes, calculated over the distribution of the values for each of the input variables. Kriging is then applied repeatedly for each of the combinations of the input values specified by the Chebyshev for each of the input dimension axes. The results are compiled into the output distribution and displayed in the uncertainty panel.

Fig. 11(b) shows the input and output distribution for two selected points from the Parallel Coordinates plot in fig. 11(a). In 11(b), the input distributions are overlaid for both selected datapoints at the top of the panel for 5 input dimensions (BumpAmplitude, α , X(Shock), DragCount, and L/D), whereas the output distributions (NC_cp, Price) are presented below separately for each datapoint.

D. Kriging Validation

The Kriging-based interpolation in CAM is the basis for generating new data points as well as quantifying uncertainty. Our implementation is supported by the Smile Machine Learning library [36]. To confirm the approach and its implementation's effectiveness and accuracy, beyond our Aircraft Wing Design dataset, we also conducted a validation on independent benchmark obtained from the Machine Learning Repository of the University of California, Irvine [37]. We arbitrarily chose three multivariate datasets of varying sizes from different disciplines comprised of real and integer values:

- **Stock Portfolio Performance** Dataset for the weighted scoring stock selection models [38] containing the normalised investment performance indicators (such as total risk or annual value) derived from a set of weights applied to stock-picking concepts (e.g. market value, systematic risk, etc.).
- **Energy Efficiency** This dataset contains the results of an analysis of different building shapes [39] which calculates two real-valued responses: heating and cooling loads of the building specifications. The inputs describe the shape of the building using such parameters as the total surface

area, relative compactness, height, orientation, as well as wall and roof area.

- **Airfoil Self-Noise** A NASA datasets containing instances of different sizes of NACA 0012 airfoils tested in a wind tunnel to estimate the sound pressure levels [40]. The calculations are based on five parameters: frequency, angle of attack, chord length, wind speed, and suction side displacement thickness.

The details of the datasets, including the Aircraft Wing Design, are described in Table I, which contains the number of instances in each dataset as well as the number of input parameters and output variables.

TABLE I
BENCHMARK DATASETS OVERVIEW.

Dataset Name	Instances	Inputs	Outputs
Stock Portfolio	63	6	3
Energy Efficiency	768	8	2
Airfoil Self-Noise	1503	5	1
Aircraft Wing Design (Aero-Manufacturing Performance)	3060	18	1

Cross Validation We employ the K-fold cross validation to verify our approach and confirm the accuracy of its predictions. More specifically, we use the leave-one-out (LOO) variant of the validation approach where $K = N$, i.e. the size of the dataset. A single sample is selected from the dataset and the interpolation model is generated using the remaining portion of the dataset. The inputs from the selected sample are then used to predict the outputs via interpolation. The value of the real output from the reserved sample is then compared with the predicted output value. This process is repeated for each instance in the dataset to obtain an overall measure of the accuracy of the predictions. To quantify it, we use the Mean Absolute Percentage Error (MAPE) [41], a well-known and widely used prediction accuracy measure.

Table II shows the MAPE-based prediction accuracy measurements for the Aircraft Wing Design study, as well as the arbitrarily chosen benchmark test datasets. As can be seen, the accuracy measurements for the benchmark datasets are very encouraging and are well within the 10% margin of error.

Most importantly, cross validation of the Aircraft Wing Design dataset, which is the central point of our study, revealed a prediction accuracy of 99.31%. The validation was set up to predict the price of manufacturing each given wing design specification, based on both the aerodynamic characteristics as well as the manufacturing process and parameters.

V. CONCLUSION AND FUTURE WORK

We presented an interactive and enhanced Parallel Coordinates plot implemented in the Cambridge Advanced Modeller.

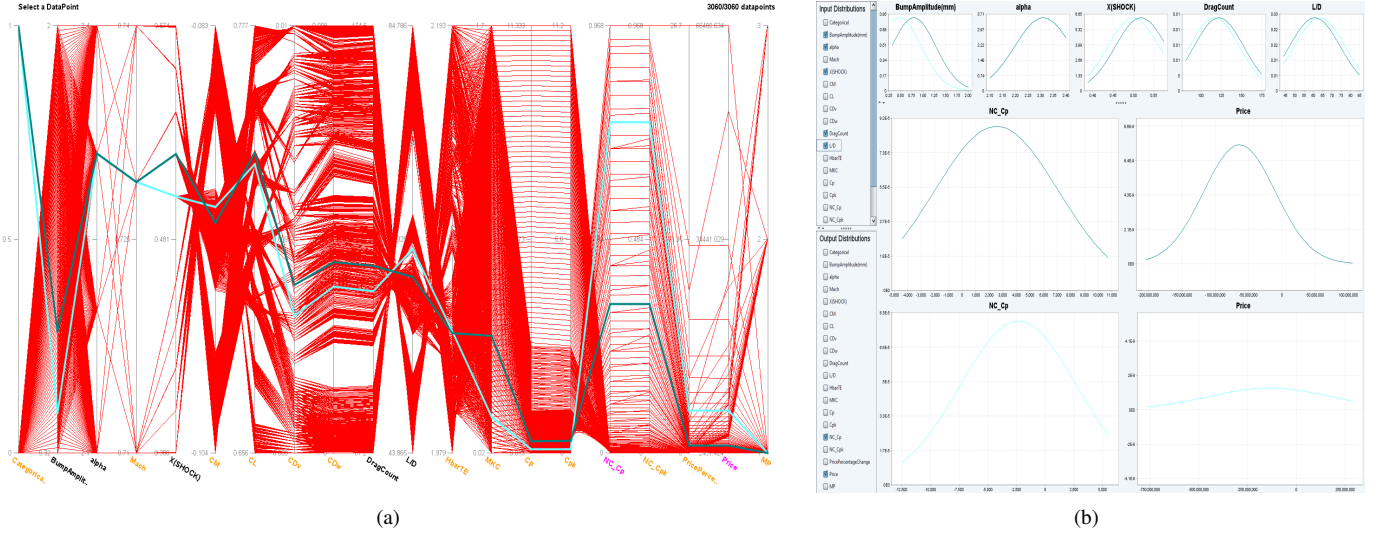


Fig. 11. Input and output value distributions for the selected datapoint

TABLE II
PREDICTION ACCURACY RESULTS FROM CROSS VALIDATION OF TEST DATASETS.

Dataset Name	Prediction Accuracy
Stock Portfolio	93.089%
Energy Efficiency	97.877%
Airfoil Self-Noise	96.626%
Aircraft Wing Design	99.309%

Our software is particularly geared towards the shift from traditional displays to high-resolution touch-screen devices. Furthermore, we developed the Parallel Coordinates to be an integral part of large engineering projects alongside other software packages. To the best of our knowledge, this is the first Parallel Coordinates implementation with embedded Machine Learning and Uncertainty Quantification capabilities for in-depth analysis. Furthermore, the interactive Parallel Coordinates toolkit is the only interactive visualization embedded in engineering design modeling, simulation, and analysis framework. The extended functionality is currently being tested, and will be included in subsequent public releases of the Cambridge Advanced Modeller (<http://www-edc.eng.cam.ac.uk/cam/>).

In the near future, we plan to improve the Uncertainty Quantification methods to improve the accuracy of analysis and further reduce risks associated with both input and output variables. In particular we are interested in implementing Polynomial Chaos and Stochastic Collocation approaches.

In addition, we intend to record the interval and scatter plot selections of the user and reflect these on the updated dataset. In this way, the engineer decision maker will be able to assess whether earlier decisions and preferences are still valid. With this functionality, we expect to reveal also the complexities in

the decision making process and not only in the design space of a product, or a system. This characteristic will be very valuable in modern engineering industry where model based analysis is a routine and the communication between different design teams and experts relies entirely on the exchange of digital information. To strengthen this characteristic even more the data exchange is following the MoSSEC standard [42]. The authors will present the benefits of such functionality in future publications.

Finally, our future development plans are also focused on adapting the Parallel Coordinates module to work more closely alongside other system design software. We will therefore continue to develop features which facilitate two-way communication between software packages to reduce duplicate effort throughout the project and ensure the exchange of accurate data between the design tools. Our future developments will continue to be compliant with and specifically target multi-gesture touch-screen devices.

ACKNOWLEDGMENT

The research leading to these results has received funding from the Aerospace Technology Institute (ATI) in the UK, under the Advanced Product Concept Analysis Environment (APROCON) project (Ref Num 113092). The authors also wish to thank the Engineering and Physical Sciences Research Council (EPSRC) for their financial support of this work as part of the AGILE Project (grant number EP/S003088/1).

REFERENCES

- [1] S. John Walker, "Big data: A revolution that will transform how we live, work, and think," 2014.
- [2] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popović *et al.*, "Predicting protein structures with a multiplayer online game," *Nature*, vol. 466, no. 7307, pp. 756–760, 2010.
- [3] F. Khatib, F. DiMaio, S. Cooper, M. Kazmierczyk, M. Gilski, S. Krzywda, H. Zabranska, I. Pichova, J. Thompson, Z. Popović *et al.*, "Crystal structure of a monomeric retroviral protease solved by protein folding game players," *Nature structural & molecular biology*, vol. 18, no. 10, pp. 1175–1177, 2011.
- [4] W. Piotrowski, M. Mörtens, D. Izzo, and D. Hennes, "Space hopper: a serious game crowdsourcing the design of interplanetary trajectories," *Acta Futura*, vol. 9, pp. 93–98, 2014.
- [5] J. S. Kim, M. J. Greene, A. Zlateski, K. Lee, M. Richardson, S. C. Turaga, M. Purcaro, M. Balkam, A. Robinson, B. F. Behabadi *et al.*, "Space-time wiring specificity supports direction selectivity in the retina," *Nature*, vol. 509, no. 7500, pp. 331–336, 2014.
- [6] A. Kawrykow, G. Roumanis, A. Kam, D. Kwak, C. Leung, C. Wu, E. Zarour, L. Sarmenta, M. Blanchette, J. Waldispühl *et al.*, "Phylo: a citizen science approach for improving multiple sequence alignment," *PLoS one*, vol. 7, no. 3, p. e31362, 2012.
- [7] J. Lee, W. Kladwang, M. Lee, D. Cantu, M. Azizyan, H. Kim, A. Limpachcher, S. Gaikwad, S. Yoon, A. Treuille *et al.*, "Rna design rules from a massive open laboratory," *Proceedings of the National Academy of Sciences*, vol. 111, no. 6, pp. 2122–2127, 2014.
- [8] A. Inselberg, "The plane with parallel coordinates," *The visual computer*, vol. 1, no. 2, pp. 69–91, 1985.
- [9] E. J. Wegman, "Hyperdimensional data analysis using parallel coordinates," *Journal of the American Statistical Association*, vol. 85, no. 411, pp. 664–675, 1990.
- [10] T. Kipouros, A. Inselberg, G. T. Parks, and A. M. Savill, "Parallel coordinates in computational engineering design," in *9th AIAA Multi-Disciplinary Design Optimization Specialist Conference*. AIAA, 2013, pp. AIAA–2013–1750.
- [11] T. Kipouros and A. Lewis, "Understanding the complexities in turbomachinery design by visualising the discontinuous areas in the design parameter space," in *OPT-i: International Conference on Engineering and Applied Sciences Optimization*, 2014, p. Paper 3090.
- [12] T. Kipouros and O. Isaksson, "Integrating value assessment into the computational engineering design cycle," in *OPT-i: International Conference on Engineering and Applied Sciences Optimization*, 2014, p. Paper 3620.
- [13] —, "Visual Analytics for Evaluation of Value Impact in Engineering Design," in *ECCOMAS Congress*, 2016, p. Paper 9111.
- [14] D. Wynn, D. Wyatt, S. Nair, and P. Clarkson, "An introduction to the cambridge advanced modeller," *Modelling and Management of Engineering Processes*, 2010.
- [15] A. Inselberg and B. Dimsdale, "Parallel coordinates: a tool for visualizing multi-dimensional geometry," in *Proceedings of the 1st Conference on Visualization'90*. IEEE Computer Society Press, 1990, pp. 361–378.
- [16] A. Inselberg, "Parallel coordinates: Visual multidimensional geometry and its applications," 2009.
- [17] J. Heinrich and D. Weiskopf, "State of the art of parallel coordinates," in *Eurographics (STARs)*, 2013, pp. 95–116.
- [18] R. A. Becker and W. S. Cleveland, "Brushing scatterplots," *Technometrics*, vol. 29, no. 2, pp. 127–142, 1987.
- [19] H. Hauser, F. Ledermann, and H. Doleisch, "Angular brushing of extended parallel coordinates," in *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*. IEEE, 2002, pp. 127–130.
- [20] M. R. Berthold and L. O. Hall, "Visualizing fuzzy points in parallel coordinates," *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 3, pp. 369–374, 2003.
- [21] R. Roberts, R. S. Laramée, G. A. Smith, P. Brookes, and T. D'Cruze, "Smart brushing for parallel coordinates," *IEEE Transactions on Visualization and Computer Graphics*, 2018.
- [22] Y. Zhonghua and W. Lingda, "3d-parallel coordinates: Visualization for time varying multidimensional data," in *Software Engineering and Service Science (ICSESS), 2016 7th IEEE International Conference on*. IEEE, 2016, pp. 655–658.
- [23] J. J. Miller and E. J. Wegman, "Construction of line densities for parallel coordinate plots," in *IN COMPUTING AND GRAPHICS IN STATISTICS*. Citeseer, 1991.
- [24] E. J. Wegman and Q. Luo, "High dimensional clustering using parallel coordinates and the grand tour," *Classification and Knowledge Organization*, pp. 93–102, 1997.
- [25] H. Zhou, X. Yuan, H. Qu, W. Cui, and B. Chen, "Visual clustering in parallel coordinates," in *Computer Graphics Forum*, vol. 27, no. 3. Wiley Online Library, 2008, pp. 1047–1054.
- [26] J. Johansson, P. Ljung, M. Jern, and M. Cooper, "Revealing structure within clustered parallel coordinates displays," in *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*. IEEE, 2005, pp. 125–132.
- [27] J. Heinrich and D. Weiskopf, "Continuous parallel coordinates," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1531–1538, 2009.
- [28] G. Andrienko and N. Andrienko, "Blending aggregation and selection: Adapting parallel coordinates for the visualization of large datasets," *The Cartographic Journal*, vol. 42, no. 1, pp. 49–60, 2005.
- [29] G. Ellis and A. Dix, "Enabling automatic clutter reduction in parallel coordinate plots," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 717–724, 2006.
- [30] M. Bostock, V. Ogievetsky, and J. Heer, "D³ data-driven documents," *IEEE Transactions on Visualization & Computer Graphics*, no. 12, pp. 2301–2309, 2011.
- [31] E. de Rochefort. (2017) XDAT. [Online]. Available: <http://www.xdat.org/index.php?ref=manual>
- [32] T. Avidan and S. Avidan, "Parallax—a data mining tool based on parallel coordinates," *Computational Statistics*, vol. 14, no. 1, pp. 79–89, 1999.
- [33] M. Friendly and D. Denis, "The early origins and development of the scatterplot," *Journal of the History of the Behavioral Sciences*, vol. 41, no. 2, pp. 103–130, 2005.
- [34] M. Gromada. (2018) mXparser. [Online]. Available: <http://mathparser.org>
- [35] M. A. OLIVER and R. WEBSTER, "Kriging: a method of interpolation for geographical information systems," *International Journal of Geographical Information Systems*, vol. 4, no. 3, pp. 313–332, 1990. [Online]. Available: <https://doi.org/10.1080/02693799008941549>
- [36] H. Li. Smile - Statistical Machine Intelligence and Learning Engine. [Online]. Available: <http://haifengl.github.io/smile/>
- [37] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [38] Y.-C. Liu and I.-C. Yeh, "Using mixture design and neural networks to build stock selection decision support systems," *Neural Computing and Applications*, vol. 28, no. 3, pp. 521–535, 2017.
- [39] A. Tsanas and A. Xifara, "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools," *Energy and Buildings*, vol. 49, pp. 560–567, 2012.
- [40] T. F. Brooks, D. S. Pope, and M. A. Marcolini, "Airfoil self-noise and prediction," 1989.
- [41] A. De Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean absolute percentage error for regression models," *Neurocomputing*, vol. 192, pp. 38–48, 2016.
- [42] "Modelling and simulation information in a collaborative systems engineering context," in <http://www.mossec.org/>.